BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **08006588 A**

(43) Date of publication of application: **12 . 01 . 96**

(51) Int. Cl **G10L 3/00**
G10L 3/00
G10L 3/00

(21) Application number: **06133339**

(22) Date of filing: **15 . 06 . 94**

(71) Applicant: **NIPPON TELEGR & TELEPH CORP <NTT>**

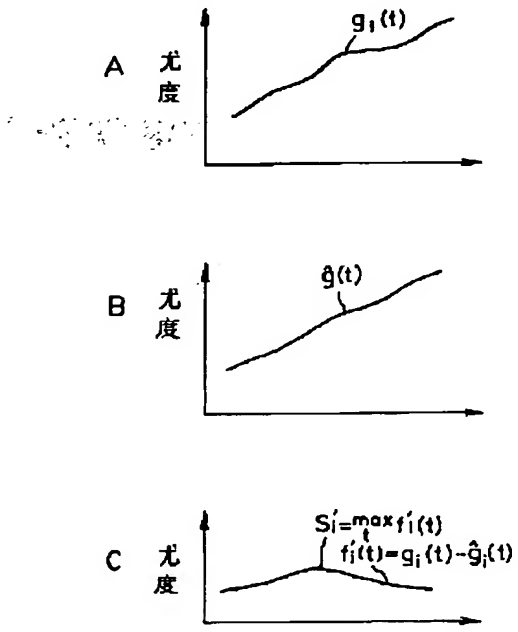
(72) Inventor: **NODA YOSHIAKI**
SAGAYAMA SHIGEKI

(54) **VOICE RECOGNITION METHOD**

(57) Abstract:

PURPOSE: To obtain an evaluation value of a hypothesis with a high precision in parallel with voice input in a beam searching method.

CONSTITUTION: Corresponding phoneme HMM and inputted voice are collated with respect to the hypothesis of the phoneme column decided by a grammar and the collating result is obtained as a liklihood function $g_i(t)$ for each hypothesis i. Then, an evaluation value is obtained from $g_i(t)$, phonemes are connected for high evaluation value hypotheses and an input voice candidate is searched. In this method, forward estimated likelihood function $g_i\Lambda(t)$ which is common to all hypotheses is obtained, $f_i'=g_i(t)-g\Lambda(t)$ is obtained, $g_i(t)$ is time normalized and let a maximum value of $f_i'(t)$ be the evaluation value of respective hypothesis i.



(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平8－6588

(43)公開日　平成８年(1996)１月12日

(51)Int.Cl.[6]	識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ１０Ｌ　3/00	５３５			
	５３１　Ｄ			
	５６１　Ｇ			

審査請求　未請求　請求項の数７　ＯＬ　（全 ８ 頁）

(21)出願番号　特願平6－133339

(22)出願日　平成６年(1994)６月15日

(71)出願人　000004226
日本電信電話株式会社
東京都新宿区西新宿三丁目19番２号

(72)発明者　野田　喜昭
東京都千代田区内幸町１丁目１番６号　日本電信電話株式会社内

(72)発明者　嵯峨山　茂樹
東京都千代田区内幸町１丁目１番６号　日本電信電話株式会社内

(74)代理人　弁理士　草野　卓

(54)【発明の名称】　音声認識方法

(57)【要約】

【目的】　ビーム探索法において、高い精度で仮説の評価値を、音声入力と並行して行うことを可能とする。

【構成】　文法により決る音素列の仮説について対応する音素ＨＭＭと入力音声とを照合し、その照合結果を各仮説ｉについて尤度関数g_i（ｔ）として得、そのg_i（ｔ）から評価値を求め高い評価値の仮説について音素を連結して入力音声候補を探索する方法において、全ての仮説に共通な前向き推定尤度関数$\hat{g}_i$（ｔ）を求め、f_i'（ｔ）＝g_i（ｔ）－$\hat{g}$（ｔ）を求めてg_i（ｔ）を時刻正規化し、このf_i'（ｔ）の最大値をそれぞれ仮説ｉの評価値とする。

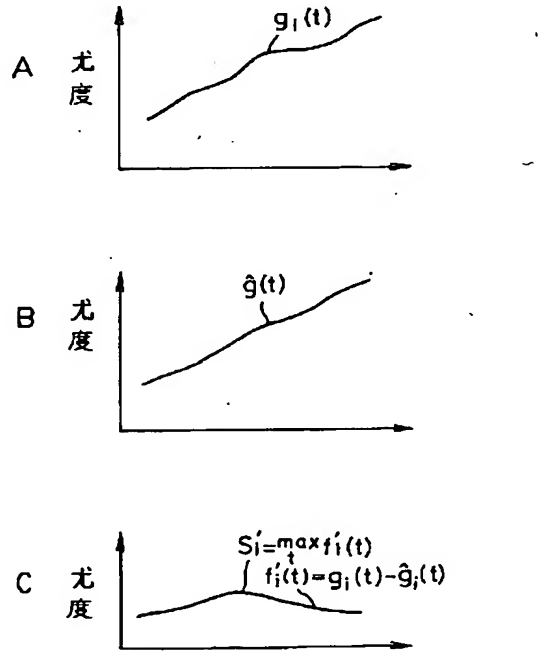


図１

【特許請求の範囲】

【請求項１】　音素のような音声単位の連結である仮説について対応する音響モデルと、入力音声とを照合し、その照合結果を尤度関数として得、その尤度関数からその仮説の評価値を求め、評価値の高い少くとも１個の仮説を残しながら入力音声に近い候補を横形探索法により探索する音声認識方法において、

全ての仮説に共通な前向き推定尤度関数を求め、

各仮説の尤度関数と上記前向き推定関数との差をとり、その差の最大値と対応する値をその仮説の評価値とすることを特徴とする音声認識方法。

【請求項２】　上記音響モデルは、隠れマルコフモデルであることを特徴とする請求項１記載の音声認識方法。

【請求項３】　上記前向き推定尤度関数を求めることは、各時刻での、全ての隠れマルコフモデルの出力確率値から最大値を選び、時刻毎にその最大値を累積することであることを特徴とする請求項２記載の音声認識方法。

【請求項４】　上記前向き推定尤度関数を求めることは、各時刻において、探索の過程で計算された隠れマルコフモデルの出力確率値の中から最大値を選び、時刻毎にその最大値を累積することであることを特徴とする請求項２記載の音声認識方法。

【請求項５】　上記前向き推定尤度関数を求めることは、音素のような音声の単位の任意の組合せと入力音声との照合によって尤度関数を求めることであることを特徴とする請求項１記載の音声認識方法。

【請求項６】　上記前向き推定尤度関数を求める際に上記音素の組合せに日本語特有の音素配列構造の制約を設けることを特徴とする請求項５記載の音声認識方法。

【請求項７】　上記前向き推定尤度関数を求めることは、探索の過程で計算された全ての仮説の尤度関数から各時刻の最大値を求め、その最大値と対応する前向き推定尤度関数を計算することを特徴とする請求項１記載の音声認識方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は、音素のような音声単位の、与えられた文法の制御に従って連結可能な数多くの各仮説について対応する音響モデルと、入力された音声とを照合し、その照合結果を尤度関数として得、その尤度関数から、その仮説の評価値を求め、評価値の高い少くとも１個の仮説を残しながら入力音声に近い候補を横形探索法により探索する音声認識方法に関する。

【０００２】

【従来の技術】図２Ａに音素を認識の単位とした音声認識処理の手順を示す。入力音声１１は、分析処理部１２により、特徴パラメータのベクトルデータ時系列に変換され、探索処理部１３により文法１６の拘束条件を用いながら、音素モデル１５との照合が行なわれる。そして、最も高い評価値を持つ音素系列が認識結果１４として出力される。

【０００３】分析処理部１２における信号処理として、よく用いられるのは、線形予測分析（Ｌｉｎｅａｒ　Ｐｒｅｄｉｃｔｉｖｅ　Ｃｏｄｉｎｇ，ＬＰＣと呼ばれる）であり、特徴パラメータとしては、ＬＰＣケプストラム、ＬＰＣデルタケプストラム、メルケプストラム、対数パワーなどがある。音素モデル１５としては確率・統計理論に基づいてモデル化された隠れマルコフモデル（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ，以後ＨＭＭ法と呼ぶ）が主流である。このＨＭＭの詳細は、例えば、社団法人電子情報通信学会編，中川聖一著『確率モデルによる音声認識』に開示されている。

【０００４】探索処理部１３は、文法で連結することが許される音素列である仮説についてその音素モデルに対して、入力音声とのもっともらしさを評価し、１つずつ仮説に音素を拡張しながら探索を進める。ここで、仮説とは、文法に示されている音素の並び順の制約に従ってつなげられた音素列のことを表し、また、仮説への音素の拡張とは、文法に従って仮説の音素列にさらに１つ音素をつなげることを意味する。

【０００５】それぞれの仮説について、１．音素列、２．トレリス計算等による、音響モデルとの照合結果である尤度関数、３．入力音声に対する仮説のもっともらしさを示す評価値、の３つの情報を記憶しておく。仮説の識別番号をｉ、時刻をｔとすると尤度関数は$g_i(t)$と表される。探索処理部１３では、まず文法によって許される１つ目の音素を仮説に拡張し、その音素に対応したＨＭＭと、分析された特徴パラメータのベクトルデータ時系列（入力音声）とを照合し、この仮説ｉの各時刻ｔの尤度$g_i(t)$を求める。ＨＭＭとの照合方法としてトレリス法、ビタービ法があり、この詳細は、例えば、社団法人電子情報通信学会編，中川聖一著『確率モデルによる音声認識』に開示されている。この尤度関数$g_i(t)$から後述する方法で仮説ｉの評価値を求め、この仮説に対し、音素列、尤度関数$g_i(t)$、評価値を記録しておく。そして、以後の音素の拡張が行なわれる毎に、その仮説に対する評価値を求めながら探索処理が進められる。また、仮説の音素列に対して、文法の制約から２種類以上の音素が拡張できる場合は、拡張できる音素の種類の数だけ元の仮説を複製し、それぞれの音素を拡張した仮説を作り、それらに対する尤度計算を行なう。このように、全ての仮説の音素列の音素数が均等となるように仮説に音素を拡張していく。文法により音素を延ばすことが出来なくなった仮説は、その音素列が文法として受理された仮説として、音素の拡張を終了する。全ての仮説で音素の拡張が出来なくなった時、文法として許される全ての音素列（仮説）に対し入力音声と照合を行なったことになり、探索処理を終える。その時の最も評価値の高い仮説の音素列また

はそれに対応する単語、文を認識結果１４として出力する。

【０００６】上記のように、探索処理において全ての仮説（音素列）の音素数を均等となるように仮説の音素を延ばす探索方法は横形探索法と呼ばれる。横形探索法を実際に行なうと、文法の許す全ての音素列に対応した仮説について計算を行なうことになり、非常に多くの仮説の計算を行なわなければならず、多くの処理時間を必要とする。このため、仮説に音素を拡張する過程で、最終的な認識結果の候補となる見込みのある仮説のみ残し、それ以外の仮説を廃棄する方法をとる場合が多い。具体的には、仮説の評価値により仮説を残すかどうかを判定する。その判定方法として仮説の評価値の高いものから順に一定個数の仮説を残す方法や、仮説の評価値のしきい値を設け、そのしきい値よりも高い仮説のみを残す方法、両者の方法の併用等が用いられる。このような横形探索法において、一定の条件により、見込みのある仮説のみを残し、それ以外の仮説を廃棄して探索を行なう方法はビーム探索法と呼ばれる。

【０００７】ビーム探索法においては、探索の途中で仮説の評価値に条件を与えて仮説の廃棄を行なうため、仮説の評価値の精度、すなわち、仮説の入力音声に対するもっともらしさを正確に評価値に反映できているか否かが、認識精度に大きな影響を与える。仮説の評価値の精度が高ければ、ビーム探索において厳しい条件で正解候補の仮説を残すことができ、処理時間を大幅に短縮できる。

【０００８】尤度関数g_i（ｔ）から仮説の評価値を求める方法について詳細に述べる。音声の始端から前向きに計算された尤度関数g_i（ｔ）は、拡張された音素までの時刻ｔでの尤度である、この尤度関数は各時刻の特徴パラメータに対する尤度をその前の時刻の尤度関数値に加えて求められる。従って、時刻が異なれば、各時刻の音素モデル内の状態遷移の出力確率を加算する回数が異なるため、時刻が異なる尤度を単純に比較することはできない。よって、尤度関数g_i（ｔ）から時刻ｔに対する最大の尤度$max\,g_i$（ｔ）を求めて、それを仮説ｉの評価値としても、時刻に対する尤度の正規化ができていないため、仮説のもっともらしさを示す値になっていない。

【０００９】以上のことを具体的に説明すると、例えば図２Ｂに示すような木構造によって表現された文法に対して、ＨＭＭを用いた探索処理を行なう場合を例とし、いま探索処理が既に第４音素までの処理を終えているとし、第５音素を拡張する場合を述べると、図２Ｂにおいては第１音素＃から４つの音素を含む仮説は、「＃ｉ_ｋ_a」，「#_ｉ_ｋ_ｉ」，「#_ｉ_ｍ_ｉ」の３種類である。ここで、“_”は音素の区切りを示す記号であり、音素＃は無音を示すものとする。

【００１０】第１音素が＃から始まり、第４音素まで拡張された一つの仮説、「#_ｉ_ｋｉ」では、図２Ｂからわかるように、第５音素として、３種類の音素ｋ，ｏ，ｍが拡張可能である。また、第１音素が＃から始まり、第４音素まで拡張されたもう１つの仮説、「#_ｉ_ｋ_ａ」は、第５音素として、２種類の音素ｍ，ｎが拡張可能である。また、仮説「#_ｉ_ｍ_ｉ」は、第４音素で完了しており、音素の拡張は行なわれない。

【００１１】音素数を一定とするビーム探索では、同じ音素数をもつ仮説に対し、仮説の評価値を求め、一定の条件で評価値の良い仮説のみを残す。ここでは、一定の条件として、評価値の高い上位２つの仮説のみを残すものとする。上で述べたように、第５音素まで拡張された仮説は、「#_ｉ_ｋ_ｉ_ｏ」，「#_ｉ_ｋ_ｉ_ｋ」，「#_ｉ_ｋ_ｉ_ｍ」，「#_ｉ_ｋ_ａ_ｍ」，「#_ｉ_ｋ_ａ_ｎ」の５種類あり、それぞれの仮説の評価値はこの順に高いとすると、上位２つの仮説である「#_ｉ_ｋ_ｉ_ｏ」と「#_ｉ_ｋ_ｉｋ」のみが次の音素を拡張できる仮説として残し、それ以外の仮説を廃棄する。

【００１２】このように、仮説に音素を拡張して、一定の条件によって残す仮説を限定し、残された仮説にさらに音素を拡張していき、全ての仮説で音素を拡張できなくなるまで、同様の処理を続ける。そして、音素を拡張できなくなった全ての完了した仮説の評価値を比較して、評価値の最も高い仮説を認識結果として、出力する。

【００１３】次に、仮説の評価値の求め方として、第４音素まで拡張された仮説「#_ｉｋ_ｉ」に音素ｏを拡張するときの、評価値の計算方法を図３Ａを用いて説明する。図は、音素列と入力音声の照合であるトレリス計算を行なって得られる尤度関数を、音素列、入力音声、尤度の３つの軸をもつ３次元の図によって示しており、図３Ａの尤度関数３１，３２に達する尤度軸と平行な直線の各長さは、各時刻の尤度の高さを示している。

【００１４】既に計算されている、仮説「#_ｉ_ｋ_ｉ」の尤度関数３１の各時刻の尤度を初期値として、トレリス計算により音素ｏの各時刻の尤度を求め、これを尤度関数３１に加えることにより、音素ｏを拡張した仮説「#_ｉ_ｋ_ｉ_ｏ」の尤度関数３２を求める。トレリス計算の計算範囲は、「#_ｉ_ｋ_ｉ」までの範囲から音素ｏの継続時間を考慮して求める。

【００１５】トレリス計算は、音響モデルを示すＨＭＭと入力音声を分析した特徴パラメータのベクトル時系列データとの照合であり、時刻ｔでＨＭＭの最終状態に到達するＨＭＭの全ての遷移に対してベクトル時系列データとの確率計算を行ない、その結果時刻ｔにおける確率値を得ることができる。ここではその確率値のｌｏｇ値である尤度を用いる。

【００１６】図３Ａにおいて曲線３３は各音素（モデル）を最も速く遷移した場合の音素列の時間経過を示

し、曲線３４は各音素（モデル）を最も長い時間かけて遷移した場合の音素列の時間経過を示す。尤度関数３１の時間軸方向の長さは音素列「＃ｉ＿ｋ＿ｉ」の継続時間と対応している。１つのＨＭＭにおいて最終状態に遷移するまでの出力確率は、それまでの状態遷移ごとにその状態の出力確率が加算され、従ってループの回数が多い程、出力確率が大となるため、尤度関数３１は、音素ｉを最も速く遷移した時刻t_1の尤度g_i（t_1）に対し、音素ｉを最も遅く遷移した時刻t_nの尤度g_i（t_n）が大きく、尤度g_i（ｔ）の各時刻での尤度は異なり、時刻の経過に従って、そのモデル内の状態遷移ごとの出力確率の加算回数が多くなり、g_i（t_n）に近ずく。このため１つの仮説についての各時刻における尤度関数を単純に比較することはできない。なんらかの方法で仮説の評価値を決める必要がある。

【００１７】そこで時刻に対する尤度の正規化を含むような仮説の評価値を求める方法として、式（１）のように音声の終端から後向きに推定した全ての仮説に共通な推定尤度関数$\hat{h}$（ｔ）を求めておき、音声の始端から前向きに計算した尤度関数g_i（ｔ）に加え、音声区間全体の推定尤度関数f_i（ｔ）を求める方法がある。この方法の詳細は、例えば「南　泰浩，山田　智一，鹿野　清宏，松岡　達雄，“番号案内を対象とした大語い連続音声認識アルゴリズム”，電子情報通信学会論文誌Ａ　Ｖｏｌ．Ｊ７７－Ａ　Ｎｏ．２，ｐｐ．１９０－１９７，１９９４」に開示されている。

【００１８】

$$f_i(t)=g_i(t)+\hat{h}(t) \qquad (1)$$

入力音声の終端は例えば図３Ａにおいて時刻t_eであり、この時刻t_eからその仮説の最後の音素より、図３Ａの例では「＃＿ｉ＿ｋ＿ｉ＿ｏ」の仮説の最も速く遷移した時刻t_1'まで、全ての仮説に共通な推定尤度関数値$\hat{h}$（t_1）を後向きに推定し、また最も遅く遷移した時刻t_n'まで、全ての仮説に共通な推定尤度関数値$\hat{h}$（t_n）を後向きに推定し、同様に時刻t_1'とt_n'との間の各時刻について後向き推定を各仮説に共通に後向き推定尤度関数$\hat{h}$（ｔ）を推定すると、この音声の終端から後向きに推定された全ての仮説に共通な推定尤度関数$\hat{h}$（ｔ）は、音声の始端から前向きに計算された尤度関数g_i（ｔ）とは逆に、図に示すように時刻に対応して尤度が単調減少している。従ってg_i（ｔ）と$\hat{h}$（ｔ）との和、つまり式（１）によって求められた音声区間全体の推定尤度関数f_i（ｔ）は、図３Ｂに示すように時刻の正規化がなされている。よって、式（２）のようにこの音声区間全体の推定尤度関数f_i（ｔ）の時刻ｔに対する最大値を求めれば、その仮説ｉのもっともらしさを示す評価値S_iを得ることができ、精度の高い評価値を得ることが出来る。

【００１９】

$$S_i=\max f_i(t) \qquad (2)$$

また、音声の終端から後向きに推定した全ての仮説に共通な推定尤度関数$\hat{h}$（ｔ）の計算方法としては、任意の音素の接続を許す文法で、音声の終端から後向きにトレリス計算を行なって求める方法がある。このようにして各仮説ｉについて評価値S_iを求め、その最も大きなもの、あるいは大きなものから複数の仮説に対して、更に音素の拡張を行うことを同様にしてゆき、拡張不能になった時の最も評価値が高い仮説を認識結果とする。

【００２０】

【発明が解決しようとする課題】しかし、上記の従来方法では、後向きの推定尤度関数$\hat{h}$（ｔ）を得るために、音声の終端から計算を行なうことになり、入力音声全体が入力されないと探索が開始できない、つまり、音声の入力と並行して探索処理を進めることが出来ない。

【００２１】音声認識において、実時間で入力される音声を実時間で認識処理し、できるだけ早い時間で認識結果が得られることは、音声認識の使いやすさを良くするものであり、実使用での音声認識に重要である。この発明は、実時間で認識処理を行なうために、音声入力と並行して探索処理を実行する仮説の評価値の計算方法を用い、しかも高精度の評価値が得られる音声認識方法を提供することにある。

【００２２】

【課題を解決するための手段】この発明によれば、ビーム探索法で尤度関数g_i（ｔ）から仮説の評価値を求める際に、音声の始端から前向きに計算された尤度関数g_i（ｔ）の時刻に対し正規化するために、音声の始端から前向きに推定した全ての仮説に共通な推定尤度関数$\hat{g}$（ｔ）を求め、音声の始端から前向きに計算した各仮説の尤度関数g_i（ｔ）からこの共通の前向き推定尤度関数$\hat{g}$（ｔ）を差し引くことにより推定尤度関数f_i'（ｔ）を得、この推定尤度関数f_i'（ｔ）は、音素列の入力音声に対する各時刻でのもっともらしさのみを含むので、このf_i'（ｔ）の最大値と対応した値を仮説の評価値として用いる。

【００２３】この方法は、音声終端からの後向き尤度関数を用いていないので、音声入力の完了を待つことなく、探索処理を並行して行なうことが出来る。

【００２４】

【実施例】以下この発明の実施例を説明する。従来と同様に入力音声を分析処理し、特徴パラメータのベクトルデータは系列に変換し、探索処理により文法の拘束条件を用いながら、ＨＭＭとの照合を、各仮説についてそれを拡張するように行い、その照合結果として各拡張音素ごとにトレリス計算により各時刻の尤度を求める。

【００２５】このトレリス計算によって得られる各時刻ｔにおけるその仮説の尤度g_i（ｔ）は、時刻ｔに対する尤度の正規化がされていない。そこでこの発明では、各仮説に共通な前向きの推定尤度関数$\hat{g}$（ｔ）を求め、式（３）のように、この仮説の尤度関数g_i（ｔ）

からｇ＾（ｔ）を差し引くことによって正規化尤度関数ｆｉ′（ｔ）を得る。前向きの推定尤度関数ｇ＾（ｔ）は正解と推定される仮説の尤度関数であって時刻ｔに対して単調に増加する。従って尤度関数ｇｉ（ｔ）が例えば図１Ａに示すように時刻ｔに対し、増加する関数であるが、前向き推定尤度関数ｇ＾（ｔ）は図１Ｂに示すよ

$$f_i'(t) = g_i(t) - \hat{g}(t) \qquad (3)$$

よって、式（４）のように、正規化尤度関数ｆｉ′（ｔ）の最大値Ｓｉ′を求めると、Ｓｉ′は仮説のもっともらしさを示している。よって、これを仮説の評価値

$$S_i' = \max f_i'(t) \qquad (4)$$

次に、前向きの推定尤度関数ｇ＾（ｔ）を求める方法について説明する。

＜前向きの推定尤度関数の計算方法１＞各音素ＨＭＭは、通常３つ程度の状態をもっており、その各状態では、複数の確率関数の重み和の出力確率分布をもっている。ここで、各時刻での特徴パラメータを全ての出力確率分布に与え、最も高い出力確率値を選択する。この出力確率値からその対数である最大尤度を求める。時刻ｔでの特徴パラメータＯτの出力確率値Ｐｉ（Ｏτ）の最大値ｍａｘＰｉ（Ｏτ）を各時刻で求め、時刻進行でのこの累積尤度を求め、それを各時刻ｔの前向きの推定尤度関数ｇ＾（ｔ）とする。式では次のように表わせる。

【００２８】

$$\hat{g}(t) = \Sigma \max P_i(O\tau) \qquad (5)$$

Σはτ＝０からｔまで

つまり式（５）は文法の拘束を外し、全てのＨＭＭの状態から何れのＨＭＭの状態へも遷移可能とし、かつその遷移確率を１として入力音声との照合をビタビ法で行なった時の各時刻での前向き最大尤度を意味しており、これをｇ＾（ｔ）とする。Ｐｉ（Ｏτ）は音声認識のためのトレリス計算の過程で可成り行われているから、その結果を利用でき、計算量が少なくて済む。

＜前向きの推定尤度関数の計算方法２＞前向きの推定尤度関数の計算方法１においては、全ての出力確率分布から得られる出力確率値の最大値から求めたが、この計算方法２では、探索処理の過程で現在までにトレリス計算によって既に計算済みの全ての出力確率分布の出力確率値の最大値から求める。このようにすると探索処理過程で文法の拘束を受けているため、これにより無関係のものが外され、しかもトレリス計算で既に計算されているためｇ＾（ｔ）のための計算をほとんど必要としない。

＜前向きの推定尤度関数の計算方法３＞横形探索法の説明で述べたように仮説に音素を拡張していき、トレリス計算を行なうことにより尤度関数を得るが、この場合、各仮説に対し、任意の音素の拡張を行なえるような文法で、つまり文法に拘束を行うことなく音素を拡張していき、得られた尤度関数を前向きの推定尤度関数とする。つまり後向き推定尤度関数ｈ＾（ｔ）と同様に文法に制約されない可能性の全ての音素配列に対する前向き推定尤度関数は最もらしい仮説に対する尤度関数に近いものとなるから、これをｇ＾（ｔ）とする。

＜前向きの推定尤度関数の計算方法４＞前向きの推定尤度関数の計算方法３においては、任意の音素の拡張を行なえるようにしたが、日本語特有の音素配列構造のみを許す制約（文法）を与えて、尤度計算を行ない、得られた尤度関数を前向きの推定尤度関数ｇ＾（ｔ）とする。日本語特有の音素配列構造を許す音素列とは、例えば「ｏ_ ｍ_ ｏ_ ｓｈ_ ｉｒ_ ｏ_ ｉ」や「ｓ_ ｕ_ ｔ_ ｏ_ ｒ_ ａ_ ｉ_ ｋ_ ｕ」というように子音の後には子音が来ないという制約を示している。「ｓ_ ｔ_ ｒ_ ａｉ_ ｋ」という音素の連鎖は英語での音素配列構造を満たしているが、日本語の音素配列構造とはなっていない。

＜前向きの推定尤度関数の計算方法５＞最終的に全ての仮説明の尤度関数の中で最大のものが正解であることがおおいから各時刻においても、全ての仮説の尤度関数中の最大のものが正解に近いと思われる。そこで探索の過程で計算された全ての仮説の尤度関数の各時刻毎の最大値を前向きの推定尤度関数とする。式で表現すると次のようになる。

【００２９】

$$\hat{g}(t) = \max g_i(t) \qquad (6)$$

上記による方法の何れかで、前向きの推定尤度関数ｇ＾（ｔ）を計算し、これを用いることにより得られる仮説の評価値を使って、ビーム探索を行なう。この仮説の評価値の精度が高いため、ビーム探索の条件を厳しくしても正確の仮説を落すことなく、探索が行なえる。また、ビーム探索の条件を厳しくすることができるため、計算すべき仮説の数を減らすことができ、探索処理量そのものを削減できる。さらに、上で述べた前向きの推定尤度関数ｇ＾（ｔ）は、各仮説の尤度関数ｇｉ（ｔ）を計算するのと同時刻までの音声データのみを必要とするため、音声の終端を待つことなく、音声の入力と並行して、探索処理を行なえる。

うに単調増加関数であって、これらの差ｆｉ′（ｔ）は図１Ｃに示すように時刻ｔに対し正規化された尤度となる。このようにｆｉ′（ｔ）は、時刻の正規化が行なわれており、時刻ｔでのその仮説のもっともらしさのみを示している。

【００２６】

とする。このようにして、時刻に対する正規化を行なった高精度の仮説の評価値を求めることができる。

【００２７】

【００３０】式（３）において、ヒューリスティック力を強くするため、つまり正解仮説を発見し易くするために、前向推定尤度関数ｇ＾（ｔ）に重みWを与え、つまりｆｉ′（ｔ）＝ｇｉ（ｔ）－Ｗｇ＾（ｔ）の演算を行うとよい。この重みWは実験的に求められる。例えば実験条件を下記に示す。

分析条件　サンプリング周波数：１２ｋＨｚ
フレーム周期：８ｍ秒
窓幅：３２ｍ秒
プリエンファシス：０．９７
特徴量：ＬＰＣケプストラム（１６次），Δケプストラム（１６次），Δ対数パワー
音響モデル　混合連続分布ＨＭＭ，状態数３，混合分布数４，対角化共分散行列
音素モデル数：５４
評価用音声データ　音素バランス２１６単語
話者：ＭＡＵ，ＭＨＴ，ＦＡＦ，ＦＳＵ
電子協提案の１００都市名
話者：男性Ａ，男性Ｂ，女性Ａ，女性Ｂ

先ず全探索を行って、最適候補（最終的に最も評価値が高くなる候補）の尤度関数を求め、先に述べた計算方法１～３と５とをそれぞれ用いて前向き推定尤度関数ｇ＾（ｔ）を求め、次に最適候補の尤度関数とｇ＾（ｔ）との距離を単位時間当りの対数尤度差とみなして、この距離が最小になるように重みWを決定する。このようにして２１６単語及び１００都市名のタスク（各タスク男性話者１名）の全ての単語に対して重みWを求めた結果は下記のようになった。

【００３１】

	２１６単語（ＭＡＵ）	１００都市名（男性Ａ）
方法１	０．９１３	０．９１４
方法２	１．０２２	１．０２４
方法３	０．９５８	０．９６２
方法５	０．９７４	０．９８１

重みWは真値と推定値の文法の制約の違いによるものと考えられ、１に近いほど両者の文法の制約が近いことを示している。

【００３２】先の実験で得られた重みWの値を用いて、２１６単語及び１００都市名を対象とした単語認識実験を行った結果を示す。重みWの値としては両タスクで推定された重みWは近い値であるから、共通に用い、計算方法１ではＷ＝０．９１、計算方法２ではＷ＝１．０２、計算方法３ではＷ＝０．９６、計算方法５ではＷ＝0.98とした。この場合の認識実験結果を、全探索を行った場合と、従来の後向き推定尤度関数を用いた場合の実験結果も合わせて図４に示す。

【００３３】この結果より、この発明によれば、従来の後向き推定尤度関数ｈ＾（ｔ）を用いる場合と同程度の認識精度が得られることが理解される。計算方法２及び５では推定尤度関数ｇ＾（ｔ）を求めるために、既に計算された出力確率あるいは尤度関数のみを用いているので、認識処理全体の処理量を低く抑えることができ、処理時間も短かい。なお図４の認識処理時間は認識処理全体の処理量に対応した時間である。なお推定尤度関数の推定精度の良さは方法１、方法３、方法２、方法５の順となった。

【００３４】上述において仮説の評価値としてｆｉ′（ｔ）の最大値を用いたが、例えばその最大値となる時刻とその前後のいくつかの時刻におけるｆｉ′（ｔ）の平均値を仮説の評価値としてもよい、つまりｆｉ′（ｔ）の最大値と対応したものを評価値とする。また上述において各１個の音素を順次連結する場合に限らず、途中で複数の連続した音素を連結する場合もある。更にこの発明は音響モデルとしてＨＭＭを用いる場合に限らず、標準パターンを用い、ＤＰパッチングにより認識する場合などにも適用できる。ＤＰマッチング等では上述における尤度は距離又は類似度と呼ばれることが多く、距離の場合は小さいほど照合での適合が良いことを示し、上述での大小関係は反対となる。

【００３５】

【発明の効果】この発明の方法の効果を以下に示す。

・各仮説の尤度関数の時刻に対応する尤度の増分を打ち消すことにより、仮説の評価値を高精度に計算することができ、認識精度を向上させることができる。

・高精度に仮説の評価値を求めることが出来るので、ビーム探索の仮説を残すための条件を厳しくすることができ、探索処理そのものの処理量を削減できる。

【００３６】・仮説の評価値を計算する際に音声区間全体のデータを必要としないので、音声入力中に並行して探索処理を行なうことができ、認識結果待ち時間を短縮できる。

【図面の簡単な説明】

【図１】Ａはある仮説ｉの尤度関数ｇｉ（ｔ）の例を示す図、Ｂは共通の前向き推定尤度関数ｇ＾（ｔ）の例を

示す図、Ｃは時刻正規化された尤度関数ｆｉ′（ｔ）の例を示す図である。

【図２】Ａは音素を認識単位とする音声認識処理の手順を示す図、Ｂは木構造に表現された音素連結の文法例を示す図である。

【図３】Ａはトレリス計算の結果得られた尤度関数の例を示す図、Ｂは時刻正規化された尤度関数ｆｉ（ｔ）の例を示す図である。

【図４】この発明方法、及び従来法による単語音声認識の実験結果を示す図である。

【図１】

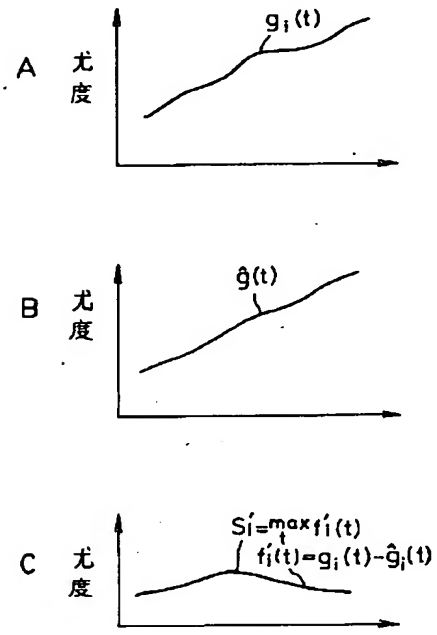


図１

【図２】

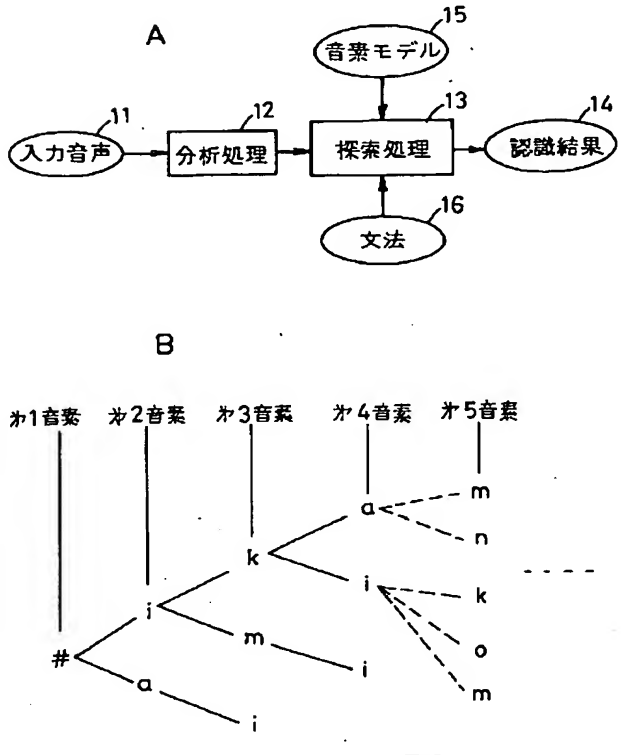


図２

【図４】

図４

Ａ

216 単語タスクでの認識率 (%)(ビーム幅 40)

	MAU	MHT	FAF	FSU	平均
全探索	94.4 (5.09)	95.4 (4.67)	93.5 (5.14)	96.8 (5.33)	95.0 (5.06)
後向き尤度	93.1 (3.03)	94.0 (2.80)	90.7 (3.12)	94.4 (3.19)	93.1 (3.03)
方法１	93.1 (2.62)	94.4 (2.44)	89.8 (2.65)	94.4 (2.76)	92.9 (2.62)
方法２	93.1 (1.83)	93.1 (1.74)	89.8 (1.83)	94.0 (1.87)	92.5 (1.82)
方法３	93.5 (3.48)	94.4 (3.20)	90.7 (3.53)	93.5 (3.65)	93.0 (3.46)
方法５	88.4 (1.78)	94.0 (1.70)	89.8 (1.80)	93.1 (1.82)	91.3 (1.78)

（括弧内は認識処理時間（秒）を示す）

Ｂ

100 都市名での認識率 (%)(ビーム幅 20)

	男性Ａ	男性Ｂ	女性Ａ	女性Ｂ	平均
全探索	94.0 (2.49)	90.0 (2.47)	88.0 (2.51)	98.0 (2.29)	92.5 (2.44)
後向き尤度	89.0 (2.67)	88.0 (2.62)	85.0 (2.64)	97.0 (2.43)	89.8 (2.59)
方法１	90.0 (2.25)	90.0 (2.22)	86.0 (2.76)	97.0 (2.65)	90.8 (2.47)
方法２	93.0 (1.22)	91.0 (1.21)	85.0 (1.20)	97.0 (1.17)	91.5 (1.19)
方法３	91.0 (2.95)	90.0 (2.90)	87.0 (2.92)	96.0 (2.67)	91.0 (2.89)
方法５	89.0 (1.16)	87.0 (1.13)	82.0 (1.14)	94.0 (1.11)	88.0 (1.14)

（括弧内は認識処理時間（秒）を示す）

【図３】

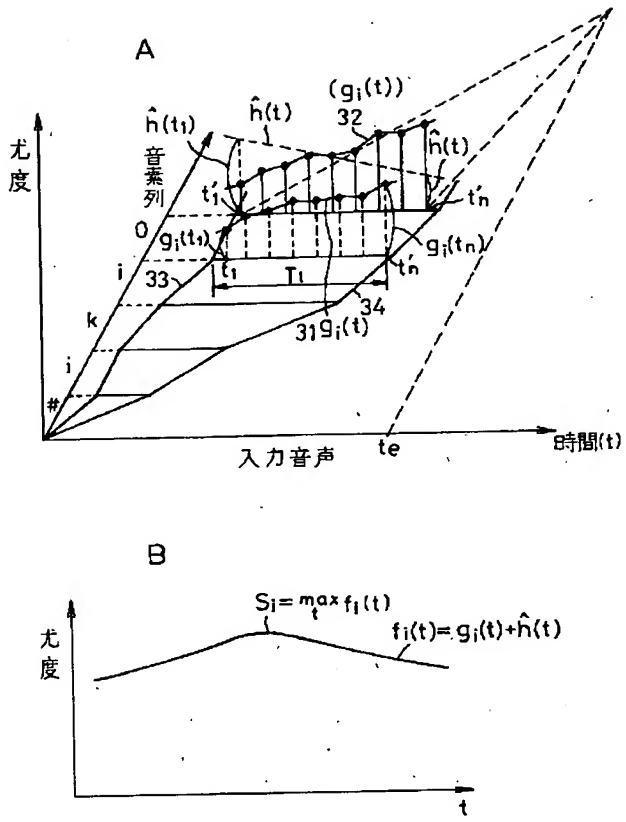


図３